

An approach to the assessment of medical problem solving : Computerised Case-based Testing

Citation for published version (APA):

Schuwirth, L. W. T. (1998). *An approach to the assessment of medical problem solving : Computerised Case-based Testing*. [Doctoral Thesis, Maastricht University]. Universiteit Maastricht. <https://doi.org/10.26481/dis.19981216ls>

Document status and date:

Published: 01/01/1998

DOI:

[10.26481/dis.19981216ls](https://doi.org/10.26481/dis.19981216ls)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 04 May. 2023

SUMMARY

When in 1974 the Maastricht University started its Problem-Based Learning curriculum, in its first faculty, the school of medicine, it was essential that a high quality assessment system was established which, on the one hand, would fit the educational approach and, on the other hand, would be useful in ensuring the quality of the graduates. A number of criteria were defined that each assessment instrument should meet: for example, congruence between assessment and education, separation of teacher and assessor role, comprehensive assessment that does not focus on certain aspects of medical competence only. In the past two decades several high-quality assessment instruments have been developed assessing mainly knowledge and skills. In the field of assessing problem-solving ability, however, a lacuna still existed. The examinations in this field, particularly those used in the clerkships varied considerably, ranging from structured orals with so-called long cases to written factual knowledge tests.

In response to this lacuna the development of Computerised Case-based Testing (CCT) was started. The idea to seek a CCT approach was founded on the many developments in psychometrics and cognitive psychology. In both disciplines it was found that problem-solving ability is not a generic skill that once mastered can be applied to any given situation, but that it is highly domain specific and idiosyncratic. This would imply that using small numbers of long and branched cases, in which the process of problem solving is measured, is not an optimal means to assess problem solving. An approach using large numbers of short cases with only a limited number of questions would be more favourable. The questions should not be aimed at general knowledge, but be focused on essential decisions. This approach is called the 'key-feature' approach. Logistically, it would be best to use computers for the administration of these examinations.

Although this approach may seem fruitful, a careful validation has not yet been performed. The main goal of this dissertation is therefore to provide evidence for the validity of CCT as a measure of problem-solving ability.

Chapter one describes the CCT method in more detail. First a description is given of the cases that are used in CCT. All case material is based on real patients, and the cases are written by the physician. This is done to optimise the authenticity of the cases. In addition, multimedia (such as pictures, sounds or moving images) are used when relevant. Cases are kept short to enable the presentation of many different topics in a test, and thus broaden the sample of topics covered.

A careful quality control system was installed to ensure the quality of the bank of cases. This entails extensive training of the authors before they start writing cases, a careful review of all case material on wording, content and relevance, and a validation of the answer keys and the authenticity by an expert panel.

Different question formats are used in CCT. The format of the question is determined by the number of realistic alternatives in real practice. Therefore, both multiple choice type and open-ended questions are used.

The main reasons to use computers in test administration are their logistical advantages, the possibility of including multimedia, and the possibility of revealing case information stepwise. Furthermore, by using a computerised database of cases, a (stratified) random test can be drawn for each student separately. A future advantage will be the use of sequential and/or adaptive testing.

A possible drawback in test design is the so-called cueing effect, which is studied in *chapter two*. This effect means that students can answer a multiple choice answer correctly by recognising the correct answer. In open-ended questions spontaneous generation is needed to produce the correct answer. It is because of this effect that multiple choice questions are assumed to be unsuitable for the testing of problem-solving ability. If this were true, it would pose a serious threat to the validity of CCT. The literature on cueing, though, is not conclusive. Results indicate that a cueing effect may exist, but that its influence on the competence being measured is negligible. By using the exactly same test twice, first presented with open-ended questions and then with multiple choice questions, a more detailed insight into the exact nature of the cueing effect could be obtained. It was found that cueing is a bidirectional effect, not only favouring the multiple choice questions, but often favouring the open-ended questions. The magnitude and direction of the cueing effect varies with the content and difficulty of the item. This led to the conclusion that the sole use of multiple choice questions would provide, to some extent, a biased estimate of the ability of the candidate.

In *chapter three* therefore, an item format specially developed for CCT was studied. It was hoped that this would have the advantage of being scorable by computers - which is impossible with open-ended questions - and would still require the student to generate the answer spontaneously. This item format was called Computerised Long Menu question (CLM). In the study CLMs were compared to hand-scored open-ended questions (OEQ) in one treatment group and to multiple choice questions (MCQ) in another. It was found that CLMs resemble open-ended questions more than they resemble multiple choice questions. Mean scores of CLMs and OEQs differed less than mean scores of CLMs and MCQs. In addition, (true) correlations between CLMs and open-ended questions were higher than those between CLMs and multiple choice questions. It was, however, also found that CLMs took more time to answer than multiple choice questions did. This may be seen as a disadvantage since it hinders broad content sampling. Therefore, CLMs can be a useful replacement for open-ended questions, but should be used only when a multiple choice question would not be valid.

In *chapter four* six research questions were studied to assess both the direct and indirect validity of CCT. For the direct validity it was studied whether students and teachers judge the content of cases and questions to be valid. Overall agreement was found that CCT poses tasks that closely resemble the problem-solving process in practice.

In addition, a multivariate generalisability procedure was used to assess the possible error variance due to the use of different question formats and different blueprint categories on the

total score. Negligible error source was found that could be attributed to either of these variables.

To assess the indirect validity it was studied whether a significant difference existed between the scores on an entrance level test and an exit level test, and whether expert scores differed from student scores on the exit level test. Finally, scores of CCT were compared to those on other measures of competence to study the pattern of correlations. It was found that a considerable increase in mean scores existed between the pre-test and the post-test, and that the experts scored even higher. Correlations between CCT and the other measures of competence were low to moderate and therefore difficult to interpret.

In summary, it was concluded that evidence for the validity of CCT emerged from the data, though the correlation matrix was not conclusive.

The next study was conducted to obtain a more direct insight into the nature of the thought processes elicited by CCT. This study is described in *chapter five*. General practitioners and students were presented with CCT cases and factual knowledge questions which were matched on content. They were asked to think aloud while solving these. This was audiotaped and typed out for further analyses. Specific indicators for analysis were defined before analysing the protocols. These were based on cognitive psychological research findings on problem solving. Indicators were word counts, the number of information units of the stem (or case) that were re-addressed after the question was read, the amount of sequentially and non-sequentially re-addressed information, and two assumed thinking steps: true-false and vectors. True-false would indicate thinking steps that considered whether something is factually true or false, whereas vector operations have a magnitude (probability) and a direction. Cases elicited thinking steps that were much more in line with the cognitive operations involved in problem solving than factual knowledge questions did. It was concluded that cases do measure problem solving in contrast to factual knowledge questions, and that the research methodology applied provides a useful addition to the 'armament' currently used in the validation of examinations.

In *chapter six* a comparison was made between students of the Maastricht University and students of the University of Groningen. All previous comparisons, using knowledge-based tests, had shown no significant differences between students of different medical schools. If with CCT differences could be found this would contribute to the external validity of CCT. Students of all year groups of both faculties were presented the same CCT test.

In the first three year groups the mean scores did not differ between schools (both groups of students of these year groups follow a similar type of curriculum). Starting from the fourth year group, however, the scores started to differ. In the final year a difference of about 8% was found in favour of the Maastricht students. In addition, the response times of the Maastricht student tended to decrease with increasing year group, whereas those of the Groningen students increased. It could be concluded from this that in the Maastricht students a higher level of problem solving takes place with increasing year groups, involving a processing of information in larger so-called 'chunks'.

It was concluded on both the proficiency scores and the response times that a difference between the student groups existed which could be attributed to differences between both curricula. CCT was, therefore, held to be sensitive to differences between curricula that cannot be detected with factual knowledge-based tests.

In *chapter seven* the experience gathered over the years of item writing for CCT is collected. Practical guidelines are given on how to write cases for authentic short case-based testing. First a general framework of a case is presented, and then strategies and pitfalls are described. The use of real patients is recommended since they provide a rich and authentic source. Considerable care must be taken to ensure a realistic and complete description in the case, both at the clinical and the contextual level. Information must be presented as 'raw' as possible; any pre-interpretation must be avoided.

The question must be aimed at essential decisions pertaining explicitly to the case. The number of questions per case must be limited to enable the administration of many different topics or problems per test. Specifically recommended is a careful review of all material by colleagues and the co-operation with other institutes to build a common database of cases.

In *chapter eight* the results are summarised and the research questions on validity of CCT are answered. Subsequently CCT is put into a broader perspective on the value of testing. Since other aspects, such as reliability, influence on learning behaviour, costs and acceptability are relevant too, these are discussed. Reliability of CCT is in range of .60 to .70 which is adequate for the setting in which CCT is used. The influence on student learning behaviour is still based on unsystematic debriefings of students. It seems to have the desired impact, although further study in this area is certainly needed. CCT is costly, with production time per case being about 2-3 person-hours. Therefore, co-operation with other medical schools in the Netherlands is being actively pursued. The acceptability is high among teachers as well as among students.

In general it is concluded that CCT provides an acceptable solution for the lacuna in the assessment system at Maastricht University.

SAMENVATTING

Toen de Universiteit Maastricht in 1974 startte met haar Probleem Gestuurd Onderwijssysteem (PGO) was deze onderwijsvorm een noviteit. Daarom was het van eminent belang dat binnen dit onderwijs een toetssysteem ontwikkeld zou worden dat aan de ene kant goed overeenstemde met de onderwijskundige uitgangspunten van PGO en anderzijds goed in staat zou zijn de kwaliteit van de afstudeerders te garanderen. Om hiertoe te kunnen komen werden een aantal uitgangspunten gedefinieerd waaraan alle toetsinstrumenten zouden moeten voldoen. Onder andere waren dit: congruentie tussen onderwijs en toetsing, scheiding van docent- en examinatorel en een toetsing die gericht was op alle aspecten van medische competentie en niet slechts op enkele aspecten. In de afgelopen twee decennia zijn verschillende toetsinstrumenten ontwikkeld die voldoen aan veel van de gedefinieerde uitgangspunten, maar voornamelijk gericht waren op kennis en vaardigheden. Op het gebied van probleem-oplossend vermogen echter bestond nog steeds een hiaat. De examens die hierop gericht waren, met name diegene die gebruikt werden voor de toetsing in de co-assistentenschappen, waren erg divers. Ze varieerden van gestructureerde mondelinge examens, waarbij lange casus gebruikt werden, tot papieren op feitenkennis gerichte toetsen.

Als antwoord op deze hiaat is Computermatige Casusgerichte Toetsing (CCT) ontstaan. De idee hiervoor is gebaseerd op de ontwikkelingen binnen het gebied van de psychometrie en de cognitieve psychologie. In beide disciplines werd aangetoond dat probleem-oplossend vermogen niet de generieke vaardigheid was, die, wanneer ze eenmaal beheerst werd, in alle voorkomende situaties toegepast zou kunnen worden. Veeleer moest probleem-oplossen gezien worden als een domein-specifieke en idiosyncratische eigenschap. Dit zou inhouden dat een aanpak waarbij kleine aantallen van lange en vertakte casus, waarbij met name het proces van probleem-oplossen gemeten wordt, niet zinvol is. Daarentegen zou een aanpak met grote aantallen korte casus met een beperkt aantal vragen per casus een betere aanpak moeten zijn. De vragen zouden dan niet gericht moeten zijn op algemene kennisaspecten, maar veeleer op beslissingen die rechtstreeks betrekking hebben op de casus. Deze aanpak wordt de 'key-feature approach' genoemd. Om logistieke redenen zou het het beste zijn een dergelijk examen computermatig af te nemen.

Op het eerste gezicht moge een dergelijke aanpak de moeite waard lijken, een zorgvuldige studie naar de validiteit heeft nog niet plaatsgevonden. Het voornaamste doel van deze dissertatie is daarom de validiteit van CCT als een meetinstrument voor probleem-oplossend vermogen vast te stellen.

In *hoofdstuk 1* wordt CCT nader beschreven. Allereerst wordt een beschrijving gegeven van de soort casus die in CCT gebruikt worden. Alle casusmateriaal wordt gebaseerd op echte patiënten, en wordt geschreven door de arts die de patiënten behandeld heeft. Dit wordt gedaan om de authenticiteit van de casus te optimaliseren. Daarnaast wordt hiervoor multimedia (bijv. foto's, geluiden en filmpjes) gebruikt, doch alleen indien dit een

emeerwaarde heeft. De casus worden kort gehouden om het mogelijk te maken vele verschillende casus per toets te gebruiken, zodat de steekproef van onderwerpen zo breed mogelijk is.

Een zorgvuldige kwaliteitscontrole procedure is opgezet om de kwaliteit van de toetsen te verzekeren. Deze bestaat onder andere uit een uitgebreide training van de auteurs voordat ze casuïstiek beginnen te produceren, een zorgvuldige review procedure, waarin alle casus beoordeeld (en verbeterd worden) op relevantie, inhoud en formulering. Daarnaast worden alle casus nog gezien door een expert-panel om de antwoordsleutel en authenticiteit te beoordelen.

Verschillende vraagtypen worden gebruikt in CCT. De vraagvorm wordt bepaald door het aantal realistische opties dat in de praktijk bestaat. Zowel open als gesloten vraagvormen worden daarom gebruikt.

De voornaamste reden om computers te gebruiken in afname en scoring van de toets zijn logistieke voordelen, de mogelijkheid om multimedia te gebruiken, en de mogelijkheid om de casusinformatie stap voor stap te onthullen. Bovendien kan, door gebruik te maken van een computer database, voor iedere student een verschillende toets getrokken worden. Een verder (toekomstig) voordeel is het gebruik van sequentiële of adaptieve toetsing.

In *hoofdstuk 2* wordt het zogenaamde cueing-effect bestudeerd. Dit effect houdt in dat studenten een multiple-choice vraag correct kunnen beantwoorden doordat ze het juiste antwoord herkennen in het rijtje alternatieven. Bij open vragen is echter het spontaan genereren van het juiste antwoord een voorwaarde om de vraag correct te kunnen beantwoorden. Met name op basis van dit effect worden multiple-choice vragen ongeschikt geacht voor de toetsing van probleem-oplossend vermogen. Als dit waar zou zijn, dan zou dat serieuze negatieve consequenties hebben voor de validiteit van CCT. De literatuur over cueing, echter, is niet geheel eenduidig. Weliswaar suggereren de resultaten dat een cueing effect bestaat, maar ook dat de invloed van de vraagvorm op de te meten competentie beperkt is.

Door dezelfde toets twee maal aan proefpersonen voor te leggen, eerst met open vragen en vervolgens met gesloten vragen, kon een duidelijker beeld verkregen worden over de aard van het cueing effect. Er werd gevonden dat cueing een bidirectioneel effect is, dat niet alleen de gesloten vragen bevoordeelt, maar soms ook de open vragen. De grootte en richting van het effect varieert met de inhoud en moeilijkheidsgraad van het item. Indien CCT daarom alleen uit multiple choice vragen zou bestaan, zou de inschatting van de competentie van de studenten niet geheel accuraat zijn, hoewel de inschattingsfout beperkt zou zijn.

Daarom is in *hoofdstuk 3* een speciaal voor CCT ontwikkelde vraagvorm bestudeerd. Van deze vraagvorm werd gehoopt dat ze het voordeel zou hebben zowel per computer scorebaar te zijn (wat niet mogelijk is met open vragen), en spontane generatie van het antwoord te vereisen. Deze vraagvorm is Computerised Long Menu (CLM) genoemd. In de studie van hoofdstuk 3 is deze vraagvorm vergeleken met handgescoorde open vragen in één onderzoeksgroep en met multiple choice vragen in een andere onderzoeksgroep.

Er werd gevonden dat CLM-vragen meer op open vragen lijken dan op multiple choice vragen. De verschillen in gemiddelde scores waren kleiner en de correlaties tussen CLM's en open vragen waren hoger dan die tussen CLM's en multiple choice vragen. Er werd echter ook gevonden dat de beantwoording van CLM's meer tijd kost dan multiple choice vragen.

Dat moet gezien worden als een nadeel daar hierdoor minder van deze vragen per uur toetstijd bevraagd kunnen worden, hetgeen de breedte van de steekproef kleiner maakt. Concluderend kan gesteld worden dat CLM's een goede vervanging kunnen zijn voor open vragen, maar dat ze alleen gebruikt moeten worden als het aangewezen is.

Hoofdstuk 4 bestaat uit een zestal deelstudies. In dit hoofdstuk wordt geprobeerd een indruk te krijgen van de directe en indirecte validiteit van CCT.

Voor de deelstudies naar de indirecte validiteit zijn data gebruikt van een CCT toets die co-assistenten voorgelegd is aan het begin van hun co-assistentenschap en van een andere CCT toets aan het einde van hun co-assistentenschap. Ook zijn scores van experts verzameld. Daarnaast zijn de scores van de studenten op de eindtoets vergeleken met scores die zij behaald hadden op andere competentiedomeinen gerichte examens. Er werd een forse toename in scores tussen de begin- en eindtoets (CCT) gevonden; experts scoren nog hoger dan studenten op hun eindtoets. De correlaties tussen CCT en de andere examens zijn in het algemeen matig tot laag en daarom niet eenduidig te interpreteren.

Om een indruk te verkrijgen van de directe validiteit zijn studenten en experts gevraagd hun mening te geven over de authenticiteit en validiteit van CCT als een toets voor probleem-oplossen. Er was een hoge mate van overeenstemming onder de ondervraagden dat het oplossen van CCT casus grote overeenkomst vertoont met het oplossen van patiëntencasus in de praktijk. Ook is een multivariate generaliseerbaarheidsanalyse gedaan om te toetsen of er een mogelijke errorbron toegeschreven kan worden aan het gebruik van verschillende vraagvormen en de verschillende blauwdrukcategorieën in één toets. Geen storende errorbron kon gevonden worden die toegeschreven zou kunnen worden aan een van beide. Algemeen is geconcludeerd dat er weliswaar bewijzen voor de validiteit aanwezig zijn, hoewel de correlaties weinigzeggend zijn. Wat betreft de directe validiteit zijn de resultaten geheel positief.

In een volgende studie is geprobeerd een nog directer inzicht te verkrijgen in de denkprocessen die door CCT casus in gang worden gezet. Deze studie wordt beschreven in *hoofdstuk 5*. Huisartsen en vijfdejaars studenten geneeskunde hebben CCT casus en inhoudelijk gematchte feitenkennisvragen voorgelegd gekregen. Hen is gevraagd hardop te denken terwijl ze de casus en vragen oplosten. Dit is op cassette opgenomen en uitgetypt voor verdere analyse. Specifieke indicatoren voor analyse zijn gedefinieerd alvorens het materiaal geanalyseerd is. Deze indicatoren zijn gebaseerd op de bevindingen uit de cognitieve psychologie met betrekking tot probleem-oplossen. Als indicatoren zijn gebruikt: lengte van de protocollen, herlezen aantallen informatie-eenheden van de stam of casus nadat de vraag beantwoord is, de hoeveelheid sequentieel en niet-sequentieel herlezen informatie en twee geponeerde denkstappen: 'true-false' en 'vector'. 'True-false' stappen zijn denkstappen waarbij alleen maar afgevraagd wordt of een gegeven (geheel) waar of onwaar is, en 'vectoren' zijn denkstappen die een richting en een grootte (probabiliteit) hebben. De denkstappen die door de casus uitgelokt werden kwamen veel meer overeen met denkstappen die een rol spelen bij probleem-oplossen dan de denkstappen die door de feitenkennisvragen opgewekt werden. Er is derhalve geconcludeerd dat CCT casus meer probleem-oplossen meten dan feitenkennisvragen, en dat de hier gehanteerde onderzoeksmethodologie een waardevolle aanvulling kan zijn op het tot op heden gebruikte valideringsinstrumentarium van toetsinstrumenten.

In *hoofdstuk 6* wordt een vergelijking beschreven tussen studenten van de universiteit van Groningen en die van Maastricht. In Groningen heeft enige jaren geleden een curriculumherziening naar een probleem gestuurde onderwijsvorm plaatsgevonden. De Groningse studenten in de eerste drie jaargroepen volgden reeds het nieuwe, terwijl de studenten van de laatste drie jaargroepen nog het oude (traditionele) curriculum volgden.

In het verleden hebben studies waarbij studenten van verschillende type curricula werden vergeleken echter nooit enige significante verschillen op afstudeerniveau laten zien. Als daarom met CCT wel verschillen te zien zouden zijn zou dit een aanwijzing zijn voor een externe validiteit van CCT, daar op grond van het curriculumverschil met name verschillen in probleem-oplossen te verwachten zijn.

In de eerste drie jaargroepen verschillen de scores niet tussen beide universiteiten. Vanaf het vierde jaar beginnen de scores echter te verschillen. In het laatste jaar wordt zelfs een verschil van 8% gezien, ten faveure van de Maastrichtse studenten. Ook werd gevonden dat de responstijden van de Maastrichtse studenten de neiging hebben te dalen bij toenemende jaargroep, terwijl die van de Groningse studenten juist stijgen. Een mogelijk conclusie is dat de Maastrichtse studenten bij het stijgen van de jaargroep in staat zijn de informatie in steeds grotere eenheden (zogenaamde 'chunks') te verwerken. Dit kan gezien worden als een redelijk vergevorderde stap in het proces van expertise ontwikkeling. Bij de Groningse studenten zouden daarentegen juist uitgebreidere analytische processen plaatsvinden bij het toenemen van de jaargroep, hetgeen op een vroeger stadium van expertise ontwikkeling duidt.

Er is zowel op basis van de scorepatronen als van de responstijden geconcludeerd dat er een verschil tussen beide groepen bestaat dat kan worden toegeschreven aan aspecten van het curriculum. Er is daarom geconcludeerd dat CCT in staat is verschillen tussen curricula te detecteren, die niet gevonden kunnen worden met feitenkennisvragen.

Hoofdstuk 7 bevat de ervaringen met het schrijven van casus die in de afgelopen jaren zijn verzameld. Er worden praktische richtlijnen gegeven voor het schrijven van korte casus voor authentieke toetsing. Allereerst wordt een soort algemeen raamwerk van een casus gepresenteerd, en vervolgens worden tips en mogelijke valkuilen beschreven. Het wordt aanbevolen gebruik te maken van echte patiënten, omdat ze een rijke en authentieke bron vormen. Veel aandacht moet besteed worden aan een realistische en complete beschrijving van de casus, zowel voor wat betreft medisch technische als contextuele informatie. De informatie moet 'ruw' aangeboden worden, pre-interpretatie door de auteur moet vermeden worden.

De vragen moeten gericht zijn op essentiële beslissingen die expliciet betrekking hebben op de casus. Er moeten niet te veel vragen per casus gesteld worden, zodat per toets vele verschillende onderwerpen aan bod kunnen komen. Er wordt speciaal op gewezen dat een zorgvuldige review van alle materiaal moet plaatsvinden (door collega's), en dat samenwerking met andere instituten om tot een gemeenschappelijke database te komen sterk kostenbesparend kan zijn.

In *hoofdstuk 8* worden de resultaten tot slot samengevat, en worden de researchvragen met betrekking tot de validiteit beantwoord. Vervolgens wordt CCT in een breder perspectief over de waarde van toetsing geplaatst. Daar ook andere aspecten van een toets dan validiteit belangrijk zijn, zoals betrouwbaarheid, invloed op leergedrag, kosten en acceptatie, worden deze kort besproken. De betrouwbaarheid van CCT is in de orde van grootte van .60 - .70

hetgeen voor de setting waarin CCT nu gebruikt wordt voldoende is. De invloed op student leergedrag kan nog steeds slechts gebaseerd worden op een beperkt aantal niet-systematische interviews. Hieruit blijkt weliswaar dat CCT de gewenste invloed heeft, maar op dit gebied is zeker nog nadere studie nodig. CCT is duur, de productietijd voor één casus ligt in de orde van grootte van 2-3 persoon-uren. Daarom bestaat een actief beleid om te proberen een samenwerking met andere medische faculteiten in Nederland te starten. De mate van acceptatie van de CCT-methode is hoog, zowel onder docenten als onder studenten. In het algemeen kan gesteld worden dat CCT een acceptabele oplossing is voor de in de inleiding geconstateerde hiaat in het toetsstelsel van de Universiteit van Maastricht.